```
******************************************************************************
MPSRCH (TM)
******************************************************************************
```


MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 10:46:26 2000; MasPar time 4.61 Seconds
308.548 Million cell updates/sec

Tabular output not generated.

Title: >US-09-331-631-7
Description: (81-140) from US09331631.pep (3 of 3)
Perfect Score: 471
Sequence: 1 LQRQYQQCQGRCQEQQQGQR..........HENYHNHKKNRSEEEEGQQR 60

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:geneseqp

Statistics: Mean 24.413; Variance 96.857; scale 0.252

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	471	100.0	525	1	W62831	Theobroma cacao antimi	1.72e-38
2	471	100.0	566	1	R20181	Sequence encoded by 67	1.72e-38
3	181	38.4	590	1	W62832	Gossypium hirsutum ant	7.42e-09
4	177	37.6	666	1	W62829	Macadamia integrifolia	1.82e-08
5	170	36.1	666	1	W62828	Macadamia integrifolia	8.66e-08
6	163	34.6	625	1	W62830	Macadamia integrifolia	4.10e-07
7	117	24.8	28	1	W62841	Stenocarpus sinuatus a	8.49e-03
8	102	21.7	395	1	W03474	Mouse SRY-related prot	1.89e-01
9	100	21.2	637	1	W62837	Hordeum vulgare antimi	2.84e-01
10	99	21.0	521	1	W74802	Human secreted protein	3.47e-01
11	96	20.4	86	1	W95073	GST-HD fusion protein	6.36e-01
12	96	20.4	86	1	W95078	GST-HD fusion protein	6.36e-01
13	96	20.4	94	1	W95075	GST-HD fusion protein	6.36e-01
14	96	20.4	94	1	W95080	GST-HD fusion protein	6.36e-01
15	96	20.4	482	1	Y07067	Renal cancer associate	6.36e-01
16	95	20.2	303	1	R60054	Dirofilaria immitis pa	7.78e-01
17	93	19.7	407	1	P94879	ORF incorporated withi	1.16e+00
18	93	19.7	444	1	W90340	G. max truncated SBP1	1.16e+00
19	93	19.7	524	1	W90339	G. max SBP1 protein.	1.16e+00
20	90	19.1	919	1	P93109	Human androgen recepto	2.10e+00
21	90	19.1	919	1	W14783	Androgen receptor.	2.10e+00
22	88	18.7	593	1	W62835	Zea mays antimicrobial	3.12e+00
23	87	18.5	96	1	W99838	HIV C76S protein seque	3.80e+00
24	87	18.5	409	1	W90342	G. max truncated SBP2	3.80e+00
25	87	18.5	489	1	W90341	G. max SBP2 protein.	3.80e+00
26	87	18.5	919	1	P90996	Human androgen recepto	3.80e+00
27	86	18.3	33	1	W62836	Zea mays antimicrobial	4.62e+00
28	86	18.3	96	1	W99831	HIV E21,24P protein se	4.62e+00
29	86	18.3	371	1	W73369	Epitope tagged TBP pro	4.62e+00
30	86	18.3	918	1	R12223	Human androgen recepto	4.62e+00
31	84	17.8	68	1	R75182	Partial peptide of hum	6.81e+00
32	84	17.8	96	1	W99836	HIV H71C protein seque	6.81e+00
33	84	17.8	131	1	R75181	Partial peptide of hum	6.81e+00
34	83	17.6	35	1	R21079	Antimicrobial maize pe	8.27e+00
35	83	17.6	1206	1	W47030	Bovine N-proteinase.	8.27e+00
36	82	17.4	96	1	W99830	HIV Vpr protein sequen	1.00e+01
37	82	17.4	428	1	R73017	Murine U2AF hetrodimer	1.00e+01
38	82	17.4	1529	1	R97985	CORK potassium channel	1.00e+01
39	81	17.2	96	1	W99833	HIV A30L protein seque	1.22e+01
40	80	17.0	69	1	R13399	Protein contg. IgA bin	1.47e+01
41	80	17.0	386	1	R04828	Arp 4.	1.47e+01
42	80	17.0	386	1	R06849	Protein Arp 4.	1.47e+01
43	80	17.0	404	1	W14909	Baker's yeast negative	1.47e+01
44	80	17.0	605	1	W62838	Glycine max antimicrob	1.47e+01
45	80	17.0	737	1	W18317	Drosophila Deltex prot	1.47e+01

ALIGNMENTS

RESULT 1
ID W62831 standard; Protein; 525 AA.
AC W62831;
DT 27-OCT-1998 (first entry)
DE Theobroma cacao antimicrobial protein.
KW antimicrobial protein; infestation; control.
OS Theobroma cacao.
PN WO9827805-A1.
PD 02-JUL-1998.
PF 22-DEC-1997; AU0874.
PR 20-DEC-1996; AU-004275.
PA (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
PI Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;
DR WPI; 98-377279/32.
PT Novel anti-microbial protein from e.g. Macadamia integrifolia -
PT useful for controlling microbial infestations of plants or mammals
PS Claim 1; Page 47-49; 96pp; English.
CC The sequence is that of an antimicrobial protein which can
CC be used to control microbial infestations in plants and mammalian
CC animals.
SQ Sequence 525 AA;

```
  Query Match            100.0%;  Score 471;  DB 1;  Length 525;
  Best Local Similarity  100.0%;  Pred. No. 1.72e-38;
  Matches   60;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Db       81 LQRQYQQCQGRCQEQQQGQREQQQCQRKCWEQYKEQERGEHENYHNHKKNRSEEEEGQQR 140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       81 LQRQYQQCQGRCQEQQQGQREQQQCQRKCWEQYKEQERGEHENYHNHKKNRSEEEEGQQR 140
```

RESULT 2
ID R20181 standard; Protein; 566 AA.
AC R20181;
DT 16-APR-1992 (first entry)
DE Sequence encoded by 67 kD T. cacao protein cDNA.
KW Cocoa; flavour; vicilin; seed storage protein.
OS Theobroma cacao.
PN WO9119801-A.
PD 26-DEC-1991.
PF 07-JUN-1991; G00914.
PR 11-JUN-1990; GB-013016.
PA (MRSC) MARS UK LTD.
PI Spencer ME, Hodge R, Deakin EA, Ashton S;
DR WPI; 92-024418/03.
DR N-PSDB; Q20377.
PT Recombinant cocoa proteins - are responsible for flavour in cocoa

```
PT   beans and produced in large quantities using yeast and bacterial
PT   expression vectors
PS   Claim 4; Fig 2; 59pp; English.
CC   The inventors claim a 67 kD and 31 kD T. cacao protein, and
CC   fragments, and encoding DNAs.  The 47 kD and 31 kD proteins are
CC   derived from the 67 kD precursor.  T. cacao protein cDNA was
CC   detected in a cDNA library prepared from immature cocoa beans RNA
CC   using a probe based on the AA sequence of a CNBr peptide common to
CC   the 47 kD and 31 kD polypeptides.  Homology searches revealed close
CC   homologies between the 67 kD polypeptide and the vicilins, which are
CC   seed storage proteins.
SQ   Sequence   566 AA;

  Query Match          100.0%;  Score 471;  DB 1;  Length 566;
  Best Local Similarity 100.0%;  Pred. No. 1.72e-38;
  Matches    60;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     81 LQRQYQQCQGRCQEQQQGQREQQQCQRKCWEQYKEQERGEHENYHNHKKNRSEEEEGQQR 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     81 LQRQYQQCQGRCQEQQQGQREQQQCQRKCWEQYKEQERGEHENYHNHKKNRSEEEEGQQR 140


RESULT    3
ID   W62832 standard; Protein; 590 AA.
AC   W62832;
DT   27-OCT-1998  (first entry)
DE   Gossypium hirsutum antimicrobial protein.
KW   antimicrobial protein; infestation; control.
OS   Gossypium hirsutum.
PN   WO9827805-A1.
PD   02-JUL-1998.
PF   22-DEC-1997; AU0874.
PR   20-DEC-1996; AU-004275.
PA   (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
PI   Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;
DR   WPI; 98-377279/32.
PT   Novel anti-microbial protein from e.g. Macadamia integrifolia -
PT   useful for controlling microbial infestations of plants or mammals
PS   Claim 1; Page 49-51; 96pp; English.
CC   The sequence is that of an antimicrobial protein which can
CC   be used to control microbial infestations in plants and mammalian
CC   animals.
SQ   Sequence   590 AA;

  Query Match           38.4%;  Score 181;  DB 1;  Length 590;
  Best Local Similarity 39.0%;  Pred. No. 7.42e-09;
  Matches    23;  Conservative   16;  Mismatches  18;  Indels   2;  Gaps   2;

Db    121 QRQFQECQQHCHQQEQRPEKKQQCVRECREKYQENPWRGEREEEAEEEETEEGEQEQSH 179
          |||:|:|| :|::|:|     ||| | |:| | |   |||:|:   :   : :|:|:::
Qy     82 QRQYQQCQGRCQEQQQGQREQQQCQRKCWEQYKEQE-RGEHEN-YHNHKKNRSEEEEGQ 138


RESULT    4
ID   W62829 standard; Protein; 666 AA.
AC   W62829;
DT   27-OCT-1998  (first entry)
DE   Macadamia integrifolia antimicrobial protein.
KW   antimicrobial protein; infestation; control.
OS   Macadamia integrifolia.
FH   Key             Location/Qualifiers
FT   Peptide         1. .28
FT                   /note= "signal peptide"
FT   Protein         29. .666
FT                   /note= "mature protein"
PN   WO9827805-A1.
PD   02-JUL-1998.
PF   22-DEC-1997; AU0874.
PR   20-DEC-1996; AU-004275.
PA   (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
PI   Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;
DR   WPI; 98-377279/32.
DR   N-PSDB; V42311.
PT   Novel anti-microbial protein from e.g. Macadamia integrifolia -
PT   useful for controlling microbial infestations of plants or mammals
PS   Claim 1; Page 39-41; 96pp; English.
CC   The sequence is that of an antimicrobial protein which can
CC   be used to control microbial infestations in plants and mammalian
CC   animals.
SQ   Sequence   666 AA;

  Query Match           37.6%;  Score 177;  DB 1;  Length 666;
  Best Local Similarity 39.0%;  Pred. No. 1.82e-08;
  Matches    23;  Conservative   19;  Mismatches  14;  Indels   3;  Gaps   3;

Db    191 QREYEDCRRRC-EQQE-PRQQYQCQRRCREQQRQHGRGG-DLINPQRGGSGRYEEGEEK 246
          ||:|::|: || |||:  |:| ||||:|:|| ::: ||  :  : ::   :  |||:::
Qy     82 QRQYQQCQGRCQEQQQGQREQQQCQRKCWEQYKEQERGEHENYHNHKKNRSEEEEGQQR 140


RESULT    5
ID   W62828 standard; Protein; 666 AA.
AC   W62828;
DT   27-OCT-1998  (first entry)
DE   Macadamia integrifolia antimicrobial protein.
KW   antimicrobial protein; infestation; control.
OS   Macadamia integrifolia.
FH   Key             Location/Qualifiers
FT   Peptide         1. .28
FT                   /note= "signal peptide"
FT   Protein         29. .666
FT                   /note= "mature protein"
PN   WO9827805-A1.
PD   02-JUL-1998.
PF   22-DEC-1997; AU0874.
PR   20-DEC-1996; AU-004275.
PA   (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
PI   Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;
DR   WPI; 98-377279/32.
DR   N-PSDB; V42310.
PT   Novel anti-microbial protein from e.g. Macadamia integrifolia -
PT   useful for controlling microbial infestations of plants or mammals
PS   Claim 1; Page 34-36; 96pp; English.
CC   The sequence is that of an antimicrobial protein which can
CC   be used to control microbial infestations in plants and mammalian
CC   animals.
SQ   Sequence   666 AA;

  Query Match           36.1%;  Score 170;  DB 1;  Length 666;
  Best Local Similarity 40.7%;  Pred. No. 8.66e-08;
  Matches    24;  Conservative   14;  Mismatches  18;  Indels   3;  Gaps   3;

Db    191 QREYEDCRRRC-EQQE-PRQQHQCQLRCREQQRQHGRGGDMMNPQRGGSGRYEEGEEEQ 247
          ||:|::|: || |||:  |:|:||| :|:|| ::: ||       :    | || | :|
Qy     82 QRQYQQCQGRCQEQQQGQREQQQCQRKCWEQYKEQERGEHENYHNHK-KNRSEEEEGQQ 139


RESULT    6
ID   W62830 standard; Protein; 625 AA.
AC   W62830;
DT   27-OCT-1998  (first entry)
DE   Macadamia integrifolia antimicrobial protein.
KW   antimicrobial protein; infestation; control.
OS   Macadamia integrifolia.
FH   Key             Location/Qualifiers
FT   Peptide         1. .28
FT                   /note= "signal peptide"
FT   Protein         29. .666
FT                   /note= "mature protein"
PN   WO9827805-A1.
PD   02-JUL-1998.
PF   22-DEC-1997; AU0874.
PR   20-DEC-1996; AU-004275.
PA   (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
PI   Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;
DR   WPI; 98-377279/32.
```

```
DR   N-PSDB; V42316.
PT   Novel anti-microbial protein from e.g. Macadamia integrifolia -
PT   useful for controlling microbial infestations of plants or mammals
PS   Claim 1; Page 43-45; 96pp; English.
CC   The sequence is that of an antimicrobial protein which can
CC   be used to control microbial infestations in plants and mammalian
CC   animals.
SQ   Sequence   625 AA;

  Query Match           34.6%;  Score 163;  DB 1;  Length 625;
  Best Local Similarity  35.6%;  Pred. No. 4.10e-07;
  Matches    21;  Conservative   15;  Mismatches  21;  Indels   2;  Gaps   1;

Db     82 QQQYEQCQKRCQRRETEPRHMQICQQRCERRYEKEKRKQQKRYEEQQ--REDEEKYEER 138
          |:||:||| ||| ::   |  | ||::|  :|  : | ::  | ::   | :||  ::|
Qy     82 QRQYQQCQGRCQEQQQGQREQQQCQRKCWEQYKEQERGEHENYHNHKKNRSEEEEGQQR 140


RESULT    7
ID   W62841 standard; Protein; 28 AA.
AC   W62841;
DT   27-OCT-1998  (first entry)
DE   Stenocarpus sinuatus antimicrobial protein.
KW   antimicrobial protein; infestation; control.
OS   Stenocarpus sinuatus.
PN   WO9827805-A1.
PD   02-JUL-1998.
PF   22-DEC-1997; AU0874.
PR   20-DEC-1996; AU-004275.
PA   (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
PI   Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;
DR   WPI; 98-377279/32.
PT   Novel anti-microbial protein from e.g. Macadamia integrifolia -
PT   useful for controlling microbial infestations of plants or mammals
PS   Claim 1; Page 66; 96pp; English.
CC   The sequence is that of an antimicrobial protein which can
CC   be used to control microbial infestations in plants and mammalian
CC   animals.
SQ   Sequence   28 AA;

  Query Match           24.8%;  Score 117;  DB 1;  Length 28;
  Best Local Similarity  60.9%;  Pred. No. 8.49e-03;
  Matches    14;  Conservative    3;  Mismatches   6;  Indels   0;  Gaps   0;

Db      5 RQQQLCQMRCQQQEKDPRQQQQC 27
          || | || |||:|:   |:||||
Qy     83 RQYQQCQGRCQEQQQGQREQQQC 105


RESULT    8
ID   W03474 standard; Protein; 395 AA.
AC   W03474;
DT   23-OCT-1996  (first entry)
DE   Mouse SRY-related protein.
KW   Mouse; SRY; primer; PCR; polymerase chain reaction; amplification; probe;
KW   HMG box; human; bovine; sex; animal; birth.
OS   Mus musculus.
PN   J08154685-A.
PD   18-JUN-1996.
PF   30-NOV-1994; 319525.
PR   30-NOV-1994; JP-319525.
PA   (KACH-) KACHIKU JUSEIRAN ISHOKU GIKUTSU KENKYUKU.
DR   WPI; 96-336575/34.
DR   N-PSDB; T33007.
PT   Bovine and mouse Sry-related DNA - useful for detecting e.g. the sex
PT   of unborn animals
PS   Claim 1; Page 10-14; 21pp; Japanese.
CC   This is the amino acid sequence of a mouse SRY-related protein.  The gene
CC   was isolated from a mouse genomic library using a cDNA fragment amplified
CC   by primers T33009-10 as a probe.  The screen isolated 4 EcoRI fragments
CC   of 2.3, 2.8, 3.5 and 1.5 kb covering the gene. Sequence analysis revealed
CC   a 240 bp HMG box sequence between bases 7154-7393.  Similarity with the
CC   human SRY HMG box sequence resulted in primers being generated to amplify
CC   the human SRY HMG box sequence for use as a probe to isolate the bovine
CC   SRY-related gene (T33008).  The mouse and bovine genes are useful for
CC   determining the sex of an animal prior to birth.
SQ   Sequence   395 AA;

  Query Match           21.7%;  Score 102;  DB 1;  Length 395;
  Best Local Similarity  25.0%;  Pred. No. 1.89e-01;
  Matches    15;  Conservative   23;  Mismatches  20;  Indels   2;  Gaps   2;

Db    208 HHQEQQFHDHHQQQQQFHDHQQQQQQQQQQQFHDHHQ-QKQQFHDHHHHQQQQQFHDHQQ 266
          ::| || : : |:||| :  ||| |:   :|: :: : : : :|:|  :: :::   :|:
Qy     82 QRQYQQCQGRCQEQQQGQREQQQCQRKCWEQYKEQERGEHENYHNHKKNRSEEE-EGQQR 140


RESULT    9
ID   W62837 standard; Protein; 637 AA.
AC   W62837;
DT   27-OCT-1998  (first entry)
DE   Hordeum vulgare antimicrobial protein.
KW   antimicrobial protein; infestation; control.
OS   Hordeum vulgare.
PN   WO9827805-A1.
PD   02-JUL-1998.
PF   22-DEC-1997; AU0874.
PR   20-DEC-1996; AU-004275.
PA   (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
PI   Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;
DR   WPI; 98-377279/32.
PT   Novel anti-microbial protein from e.g. Macadamia integrifolia -
PT   useful for controlling microbial infestations of plants or mammals
PS   Claim 1; Page 60-62; 96pp; English.
CC   The sequence is that of an antimicrobial protein which can
CC   be used to control microbial infestations in plants and mammalian
CC   animals.
SQ   Sequence   637 AA;

  Query Match           21.2%;  Score 100;  DB 1;  Length 637;
  Best Local Similarity  24.1%;  Pred. No. 2.84e-01;
  Matches    13;  Conservative   20;  Mismatches  20;  Indels   1;  Gaps   1;

Db     42 QQCVQRCR-QERPRYSHARCVQECRDDQQQHGRHEQEEEQGRGRGWHGEGEREE 94
          |||  ||: |:: :  : :| : |:::  :: | |:|: : : : :  | | ::
Qy     86 QQCQGRCQEQQQGQREQQQCQRKCWEQYKEQERGEHENYHNHKKNRSEEEEGQQ 139


RESULT   10
ID   W74802 standard; Protein; 521 AA.
AC   W74802;
DT   25-JAN-1999  (first entry)
DE   Human secreted protein encoded by gene 73 clone HSQEL25.
KW   Human; secreted protein; testis; tumour; foetal brain tissue;
KW   fusion protein; cancer; central nervous system; seizure;
KW   diagnosis; neurodegenerative disease.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Misc-difference 521
FT                   label= unknown
PN   WO9839448-A2.
PD   11-SEP-1998.
PF   06-MAR-1998; U04493.
PR   02-OCT-1997; US-061060.
PR   07-MAR-1997; US-038621.
PR   07-MAR-1997; US-040161.
PR   07-MAR-1997; US-040162.
PR   07-MAR-1997; US-040163.
PR   07-MAR-1997; US-040333.
PR   07-MAR-1997; US-040334.
PR   07-MAR-1997; US-040336.
PR   07-MAR-1997; US-040626.
PR   11-APR-1997; US-043311.
PR   11-APR-1997; US-043312.
PR   11-APR-1997; US-043313.
PR   11-APR-1997; US-043314.
```

PR 11-APR-1997; US-043568.
PR 11-APR-1997; US-043569.
PR 11-APR-1997; US-043576.
PR 11-APR-1997; US-043578.
PR 11-APR-1997; US-043580.
PR 11-APR-1997; US-043669.
PR 11-APR-1997; US-043670.
PR 11-APR-1997; US-043671.
PR 11-APR-1997; US-043672.
PR 11-APR-1997; US-043674.
PR 23-MAY-1997; US-047492.
PR 23-MAY-1997; US-047500.
PR 23-MAY-1997; US-047501.
PR 23-MAY-1997; US-047502.
PR 23-MAY-1997; US-047503.
PR 23-MAY-1997; US-047581.
PR 23-MAY-1997; US-047582.
PR 23-MAY-1997; US-047583.
PR 23-MAY-1997; US-047584.
PR 23-MAY-1997; US-047585.
PR 23-MAY-1997; US-047586.
PR 23-MAY-1997; US-047587.
PR 23-MAY-1997; US-047588.
PR 23-MAY-1997; US-047589.
PR 23-MAY-1997; US-047590.
PR 23-MAY-1997; US-047592.
PR 23-MAY-1997; US-047593.
PR 23-MAY-1997; US-047594.
PR 23-MAY-1997; US-047595.
PR 23-MAY-1997; US-047596.
PR 23-MAY-1997; US-047597.
PR 23-MAY-1997; US-047598.
PR 23-MAY-1997; US-047599.
PR 23-MAY-1997; US-047600.
PR 23-MAY-1997; US-047601.
PR 23-MAY-1997; US-047612.
PR 23-MAY-1997; US-047613.
PR 23-MAY-1997; US-047614.
PR 23-MAY-1997; US-047615.
PR 23-MAY-1997; US-047617.
PR 23-MAY-1997; US-047618.
PR 23-MAY-1997; US-047632.
PR 23-MAY-1997; US-047633.
PR 06-JUN-1997; US-048964.
PR 06-JUN-1997; US-048974.
PR 13-JUN-1997; US-049610.
PR 08-JUL-1997; US-051926.
PR 16-JUL-1997; US-052874.
PR 18-AUG-1997; US-055724.
PR 22-AUG-1997; US-056630.
PR 22-AUG-1997; US-056631.
PR 22-AUG-1997; US-056632.
PR 22-AUG-1997; US-056636.
PR 22-AUG-1997; US-056637.
PR 22-AUG-1997; US-056662.
PR 22-AUG-1997; US-056664.
PR 22-AUG-1997; US-056845.
PR 22-AUG-1997; US-056862.
PR 22-AUG-1997; US-056864.
PR 22-AUG-1997; US-056872.
PR 22-AUG-1997; US-056874.
PR 22-AUG-1997; US-056875.
PR 22-AUG-1997; US-056876.
PR 22-AUG-1997; US-056877.
PR 22-AUG-1997; US-056878.
PR 22-AUG-1997; US-056879.
PR 22-AUG-1997; US-056880.
PR 22-AUG-1997; US-056881.
PR 22-AUG-1997; US-056882.
PR 22-AUG-1997; US-056884.
PR 22-AUG-1997; US-056886.
PR 22-AUG-1997; US-056887.
PR 22-AUG-1997; US-056888.
PR 22-AUG-1997; US-056889.
PR 22-AUG-1997; US-056892.
PR 22-AUG-1997; US-056893.
PR 22-AUG-1997; US-056894.
PR 22-AUG-1997; US-056903.
PR 22-AUG-1997; US-056908.
PR 22-AUG-1997; US-056909.
PR 22-AUG-1997; US-056910.
PR 22-AUG-1997; US-056911.
PR 05-SEP-1997; US-057650.
PR 05-SEP-1997; US-057669.
PR 05-SEP-1997; US-057761.
PR 12-SEP-1997; US-058785.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,
PI Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS,
PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,
PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
DR WPI; 98-506364/43.
DR N-PSDB; V59583.
PT New isolated human genes and the secreted polypeptide(s) they encode
PT - useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
PS Claim 1; Page 583-584; 721pp; English.
CC This sequence represents a secreted human protein encoded by the nucleic
CC acid molecule designated Gene 73 from the human cDNA clone HSQEL25
CC (deposited as clone ATCC 97900 and ATCC 209046).
CC The gene can be used to generate fusion proteins by linking to the gene
CC to a human immunoglobulin Fc portion (e.g. V59502) for increasing the
CC stability of the fused protein as compared to the human protein only.
CC The invention relates to 186 novel genes and their fragments (nucleic
CC acid sequences: V59511-V59812; amino acid sequences W74731-W75026) which
CC are useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. Also, pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 186 polynucleotides, based on
CC which tissues they are most highly expressed in (see V59511 for described
CC uses).
SQ Sequence 521 AA;

Query Match 21.0%; Score 99; DB 1; Length 521;
Best Local Similarity 23.3%; Pred. No. 3.47e-01;
Matches 14; Conservative 24; Mismatches 21; Indels 1; Gaps 1;

```
Db    429 ERYRYREYAERGYERHRASREKEERHRERRHREKEETRHKSSRSNSRRRHESEEGDSHRR 488
          :| :|::   |  |:::: || :: :|  : : ||: |      :::::: ||| ::::|
Qy     82 QR-QYQQCQGRCQEQQQGQREQQQCQRKCWEQYKEQERGEHENYHNHKKNRSEEEEGQQR 140
```

RESULT 11
ID W95073 standard; Protein; 86 AA.
AC W95073;
DT 20-MAY-1999 (first entry)
DE GST-HD fusion protein GST-HD51DELP.
KW Amyloid-like fibril; protein aggregate; inhibitor; inclusion body;
KW polyglutamine expansion; Huntington's disease; Alzheimer's disease; HD;
KW Parkinson's disease; spinal; bulbar muscular atrophy; type II diabetes;
KW systemic amyloidosis; spinocerebellar ataxia; kuru; familial insomnia;
KW bovine spongiform encephalopathy; kuru; scrapie; GST-HD; fusion protein.
OS Synthetic.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Misc_difference 1
FT /note= "this residue is connected to a GST protein
FT which is not indicated in the sequence"
PN WO9906838-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; E04810.
PR 01-AUG-1997; EP-113320.
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PI Bates G, Lehrach H, Scherzinger E, Wanker E;
DR WPI; 99-153955/13.

PT Detecting amyloid-like fibrils or protein aggregates insoluble in
PT detergent or urea - from their retention on a filter, used for
PT diagnosis, particularly of diseases associated with polyglutamine
PT expansion
PS Disclosure; Fig 8; 56pp; English.
CC The invention relates to the detection of amyloid-like fibrils or protein
CC aggregates, insoluble in detergents or urea. The method comprises: (a)
CC applying material suspected of containing protein aggregates to a filter;
CC and (b) detecting retention of protein aggregates on the filter. This
CC method also helps to identify inhibitors of protein aggregates formation.
CC The method is particularly used to detect protein aggregates that are
CC indicative of disease, for assessing onset or progression of the
CC diseases. The inhibitors identified are potential therapeutic agents for
CC treating the diseases. Other applications include detection of inclusion
CC bodies in bacteria and to study kinetics of aggregate formation. Diseases
CC associated with polyglutamine expansion are particularly diagnosed, e.g.
CC Huntington's, Alzheimer's or Parkinson's diseases; spinal and bulbar
CC muscular atrophy; spinocerebellar ataxia; systemic amyloidosis; type II
CC diabetes; bovine spongiform encephalopathy; kuru; familial insomnia;
CC scrapie. The protein aggregates can now be detected simply, routinely and
CC rapidly, without requiring sophisticated equipment. The method can be
CC made quantitative, by analysing a series of dilutions, and can be
CC automated to allow many samples to be analysed on the same filter.
CC Sequences W95072-75 represent GST-HD fusion proteins.
SQ Sequence 86 AA;

Query Match 20.4%; Score 96; DB 1; Length 86;
Best Local Similarity 28.1%; Pred. No. 6.36e-01;
Matches 16; Conservative 21; Mismatches 17; Indels 3; Gaps 2;

Db 22 KSFQQQQQQQQQQQQQQQQQQQQ-QQQ--QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 75
: :|| | : |:||| |::||| |: :| :|:: ::: : : : :::: ||
Qy 83 RQYQQCQGRCQEQQQGQREQQQCQRKCWEQYKEQERGEHENYHNHKKNRSEEEEGQQ 139

RESULT 12
ID W95078 standard; Protein; 86 AA.
AC W95078;
DT 20-MAY-1999 (first entry)
DE GST-HD fusion protein GST-HD51DELP.
KW Fusion protein; amyloidogenic polypeptide; amyloid-like fibril; scrapie;
KW protein aggregate; Alzheimer's disease; CAG-repeat expansion; spinal;
KW Huntington's disease; bulbar muscular atrophy; spinocerebellar ataxia;
KW dentatorubral pallidoluysian atrophy; Creutzfeld-Jakob disease; enzyme;
KW GST-HD; HD.
OS Synthetic.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Misc_difference 1
FT /note= "this residue is connected to a GST protein
FT which is not indicated in the sequence"
PN WO9906545-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; E04811.
PR 01-AUG-1997; EP-113306.
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PI Bates G, Lehrach H, Scherzinger E, Wanker E;
DR WPI; 99-153775/13.
PT Composition containing fusion protein that includes amyloidogenic
PT peptide - able to self-assemble into fibrils or aggregates, used to
PT detect and monitor neuronal diseases, and also to screen for
PT therapeutic inhibitors
PS Disclosure; Fig 8; 62pp; English.
CC The invention relates to a composition comprising a fusion protein of (i)
CC (poly)peptide that increases solubility and/or prevents aggregation of
CC fusion protein, and (ii) amyloidogenic (poly)peptide that can self-
CC assemble into amyloid-like fibrils or protein aggregates. Host cells
CC transformed with a vector containing the nucleic acid encoding the fusion
CC protein are used for the recombinant expression of the fusion protein.
CC The composition is used to detect onset and progression of diseases
CC associated with fibrils/protein aggregates. It is potentially useful for
CC treatment of such diseases (e.g. Alzheimer's disease, scrapie or CAG-
CC repeat expansion conditions such as Huntington's disease (HD), spinal and
CC bulbar muscular atrophy, dentatorubral pallidoluysian atrophy,
CC spinocerebellar ataxia, Creutzfeld-Jakob disease). Assay methods based on
CC release of the amyloidogenic polypeptide from fusion protein have a
CC precise starting time for aggregate formation, allowing kinetic
CC measurements, and use of an enzyme for cleavage allows testing under
CC physiological conditions. Sequences W95077-80 represent GST-HD fusion
CC proteins.
SQ Sequence 86 AA;

Query Match 20.4%; Score 96; DB 1; Length 86;
Best Local Similarity 28.1%; Pred. No. 6.36e-01;
Matches 16; Conservative 21; Mismatches 17; Indels 3; Gaps 2;

Db 22 KSFQQQQQQQQQQQQQQQQQQQQ-QQQ--QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 75
: :|| | : |:||| |::||| |: :| :|:: ::: : : : :::: ||
Qy 83 RQYQQCQGRCQEQQQGQREQQQCQRKCWEQYKEQERGEHENYHNHKKNRSEEEEGQQ 139

RESULT 13
ID W95075 standard; Protein; 94 AA.
AC W95075;
DT 20-MAY-1999 (first entry)
DE GST-HD fusion protein GST-HD51DELPBio.
KW Amyloid-like fibril; protein aggregate; inhibitor; inclusion body;
KW polyglutamine expansion; Huntington's disease; Alzheimer's disease; HD;
KW Parkinson's disease; spinal; bulbar muscular atrophy; type II diabetes;
KW systemic amyloidosis; spinocerebellar ataxia; kuru; familial insomnia;
KW bovine spongiform encephalopathy; kuru; scrapie; GST-HD; fusion protein.
OS Synthetic.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Misc_difference 1
FT /note= "this residue is connected to a GST protein
FT which is not indicated in the sequence"
PN WO9906838-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; E04810.
PR 01-AUG-1997; EP-113320.
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PI Bates G, Lehrach H, Scherzinger E, Wanker E;
DR WPI; 99-153955/13.
PT Detecting amyloid-like fibrils or protein aggregates insoluble in
PT detergent or urea - from their retention on a filter, used for
PT diagnosis, particularly of diseases associated with polyglutamine
PT expansion
PS Disclosure; Fig 8; 56pp; English.
CC The invention relates to the detection of amyloid-like fibrils or protein
CC aggregates, insoluble in detergents or urea. The method comprises: (a)
CC applying material suspected of containing protein aggregates to a filter;
CC and (b) detecting retention of protein aggregates on the filter. This
CC method also helps to identify inhibitors of protein aggregates formation.
CC The method is particularly used to detect protein aggregates that are
CC indicative of disease, for assessing onset or progression of the
CC diseases. The inhibitors identified are potential therapeutic agents for
CC treating the diseases. Other applications include detection of inclusion
CC bodies in bacteria and to study kinetics of aggregate formation. Diseases
CC associated with polyglutamine expansion are particularly diagnosed, e.g.
CC Huntington's, Alzheimer's or Parkinson's diseases; spinal and bulbar
CC muscular atrophy; spinocerebellar ataxia; systemic amyloidosis; type II
CC diabetes; bovine spongiform encephalopathy; kuru; familial insomnia;
CC scrapie. The protein aggregates can now be detected simply, routinely and
CC rapidly, without requiring sophisticated equipment. The method can be
CC made quantitative, by analysing a series of dilutions, and can be
CC automated to allow many samples to be analysed on the same filter.
CC Sequences W95072-75 represent GST-HD fusion proteins.
SQ Sequence 94 AA;

Query Match 20.4%; Score 96; DB 1; Length 94;
Best Local Similarity 28.1%; Pred. No. 6.36e-01;
Matches 16; Conservative 21; Mismatches 17; Indels 3; Gaps 2;

Db 22 KSFQQQQQQQQQQQQQQQQQQQQ-QQQ--QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 75
: :|| | : |:||| |::||| |: :| :|:: ::: : : : :::: ||

```
Qy      83 RQYQQCQGRCQEQQQGQREQQQCQRKCWEQYKEQERGEHENYHNHKKNRSEEEEGQQ 139


RESULT   14
ID   W95080 standard; Protein; 94 AA.
AC   W95080;
DT   20-MAY-1999  (first entry)
DE   GST-HD fusion protein GST-HD51DELPBio.
KW   Fusion protein; amyloidogenic polypeptide; amyloid-like fibril; scrapie;
KW   protein aggregate; Alzheimer's disease; CAG-repeat expansion; spinal;
KW   Huntington's disease; bulbar muscular atrophy; spinocerebellar ataxia;
KW   dentatorubral pallidoluysian atrophy; Creutzfeld-Jakob disease; enzyme;
KW   GST-HD; HD.
OS   Synthetic.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Misc_difference 1
FT                   /note= "this residue is connected to a GST protein
FT                         which is not indicated in the sequence"
PN   WO9906545-A2.
PD   11-FEB-1999.
PF   31-JUL-1998; E04811.
PR   01-AUG-1997; EP-113306.
PA   (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PI   Bates G, Lehrach H, Scherzinger E, Wanker E;
DR   WPI; 99-153775/13.
PT   Composition containing fusion protein that includes amyloidogenic
PT   peptide - able to self-assemble into fibrils or aggregates, used to
PT   detect and monitor neuronal diseases, and also to screen for
PT   therapeutic inhibitors
PS   Disclosure; Fig 8; 62pp; English.
CC   The invention relates to a composition comprising a fusion protein of (i)
CC   (poly)peptide that increases solubility and/or prevents aggregation of
CC   fusion protein, and (ii) amyloidogenic (poly)peptide that can self-
CC   assemble into amyloid-like fibrils or protein aggregates. Host cells
CC   transformed with a vector containing the nucleic acid encoding the fusion
CC   protein are used for the recombinant expression of the fusion protein.
CC   The composition is used to detect onset and progression of diseases
CC   associated with fibrils/protein aggregates. It is potentially useful for
CC   treatment of such diseases (e.g. Alzheimer's disease, scrapie or CAG-
CC   repeat expansion conditions such as Huntington's disease (HD), spinal and
CC   bulbar muscular atrophy, dentatorubral pallidoluysian atrophy,
CC   spinocerebellar ataxia, Creutzfeld-Jakob disease). Assay methods based on
CC   release of the amyloidogenic polypeptide from fusion protein have a
CC   precise starting time for aggregate formation, allowing kinetic
CC   measurements, and use of an enzyme for cleavage allows testing under
CC   physiological conditions. Sequences W95077-80 represent GST-HD fusion
CC   proteins.
SQ   Sequence   94 AA;

  Query Match           20.4%;  Score 96;  DB 1;  Length 94;
  Best Local Similarity  28.1%;  Pred. No. 6.36e-01;
  Matches   16;  Conservative   21;  Mismatches  17;  Indels   3;  Gaps   2;

Db      22 KSFQQQQQQQQQQQQQQQQQQQQ-QQQ--QQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 75
           : :|| | : |:||| |::||| |:   :|  :|:: :::  :  :    : :::: ||
Qy      83 RQYQQCQGRCQEQQQGQREQQQCQRKCWEQYKEQERGEHENYHNHKKNRSEEEEGQQ 139


RESULT   15
ID   Y07067 standard; Protein; 482 AA.
AC   Y07067;
DT   02-JUL-1999  (first entry)
DE   Renal cancer associated antigen precursor sequence.
KW   Cancer associated antigen; diagnosis; research; treatment; human;
KW   breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW   prostate cancer.
OS   Homo sapiens.
PN   WO9904265-A2.
PD   28-JAN-1999.
PF   15-JUL-1998; U14679.
PR   22-JUN-1998; US-102322.
PR   17-JUL-1997; US-896164.
PR   10-OCT-1997; US-061599.
PR   10-OCT-1997; US-061765.
PR   10-OCT-1997; US-948705.
PR   11-OCT-1997; GB-021697.
PA   (LUDW-) LUDWIG INST CANCER RES.
PI   Chen Y, Gout I, Gure A, OHare M, Obata Y, Old LJ,
PI   Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E,
PI   Tureci O;
DR   WPI; 99-132448/11.
PT   New isolated cancer associated nucleic acids and polypeptides -
PT   isolated using sera from cancer patients, used to develop products
PT   for the diagnosis, monitoring or treatment of cancers
PS   Disclosure; Page 467-468; 787pp; English.
CC   The invention relates to a method for diagnosing a disorder characterised
CC   by expression of a human cancer associated antigen precursor coded for by
CC   a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC   biological sample isolated from a subject with an agent that specifically
CC   binds to the NAM, an expression product or a fragment of an expression
CC   product complexed with an HLA molecule; and (b) determining the
CC   interaction between the agent and the NAM or the expression product as a
CC   determination of the disorder. The products and methods can be used in
CC   the diagnosis, monitoring, research, or treatment of conditions
CC   characterised by the expression of various cancer associated antigens.
CC   The invention provides nucleic acid sequences and encoded polypeptides
CC   which are cancer associated antigen precursors expressed in human breast
CC   cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC   lung cancer.
SQ   Sequence   482 AA;

  Query Match           20.4%;  Score 96;  DB 1;  Length 482;
  Best Local Similarity  34.0%;  Pred. No. 6.36e-01;
  Matches   16;  Conservative   13;  Mismatches  17;  Indels   1;  Gaps   1;

Db      93 EEAAKKRQEEQ-ERKLKEQWEEQQRKEREEEEQKRQEKKEKEEALQK 138
           |:    |:::| :||  ||  ||:| |:|:    : :: | ||: |:
Qy      94 EQQQGQREQQQCQRKCWEQYKEQERGEHENYHNHKKNRSEEEEGQQR 140


Search completed: Sat May 13 10:46:34 2000
Job time : 8 secs.
```